Review

Cognitive psychology and the design of alarm sounds

J. Edworthy and C.S. Meredith

Department of Psychology, University of Plymouth, Plymouth, UK

Received 14 September 1993 accepted 6 December 1993

ABSTRACT
One of the less desirable features about technological advances in medicine is that individuals working with sophisticated monitoring equipment are often bombarded by warning sounds and signals. However, there are some basic cognitive psychological principles which suggest that the over-use of auditory warnings in such circumstances may be counter-productive. This review highlights some of these principles, arguing that future systems should place more emphasis upon cognitive capabilities and predispositions in their design. Upcoming alarms standards where issues such as the number of alarms, their recognizability, and the principle of 'urgency mapping' are discussed with relation to the cognitive psychology of sound.

Keywords: Alarms, cognitive psychology, warning design

INTRODUCTION
Medical alarms and monitoring sounds are often numerous, inappropriate, and confusing. This is not just hearsay, but has been demonstrated empirically (for example, Montahan et al.1). However, there is a growing number of attempts to make such systems more ergonomic and acceptable2,3,4. Currently there is also a major review of standards at a British, European and worldwide level for some types of Intensive Therapy Unit alarms, which will ultimately have a significant impact on the design of future warnings systems. This article will review some of these developments at a psychological level, drawing attention to the types of features which more ergonomic, acceptable alarm systems will need to possess in order to improve the current situation, and to fall in line with these new standards. These standards are, to some extent, based upon important psychological principles. The article focuses only on auditory alarms, although it is acknowledged that there are a number of interesting and potentially very fruitful developments in the design of visual displays, central monitoring devices and so on which could be integrated with auditory alarms. This article also does not consider the relationship between medical pathology and the alarm sounds which might accompany this development; rather, it focuses on the design of the alarm sounds themselves and not the conditions which will evoke alarms.

If a nurse or doctor is attending to some other task such as drug preparation some distance from the patient, an auditory warning will usually be the most effective signal of a developing medical situation. Alarms need not remain in their current generally badly-designed form, however. The central purpose of this article is to draw attention to a number of important psychoacoustic responses which the population as a whole demonstrate. These are relatively unchanging, unlike technology, and therefore should be borne in mind as a central rather than as a peripheral feature of the person–machine system of which alarms and their devices are an example.

ALARM USE IN INTENSIVE THERAPY UNITS AND OPERATING ROOMS
It is not unusual to find 50 or 60 alarms in a multibedded Intensive Therapy Unit. Studies have shown that a single patient can be attached to several pieces of equipment, each of which has the potential to produce a range of alarms. This can result in 20 or more alarms for a single patient5. Many of these alarms are unnecessarily loud and continuous, which results not only in irritation for the medical staff, but disturbance of the patient who, when seriously ill, has a lower tolerance for noise than the well individual6.

The first requirement of an alarm is that it is audible. Although many are unnecessarily loud, a
study\(^1\) shows that there is the distinct possibility that some alarms, and some pieces of equipment, will mask others if they are heard together. This study also shows that medical staff recognized an average of only half of the 50 possible warnings when tested for recognition. These two problems are related. If there are many warnings, then it is beyond the cognitive capacity of most people (not only medical staff) to recognize such a high number of alarms; additionally, the larger the number, the greater the chances that masking will occur, which promotes the need to create louder and louder alarms, which in turn is not ergonomic.

It is often assumed by the general public (which includes the patients) that an alarm signifies a critical medical event. One study by O’Carroll\(^7\) required nurses to record alarm soundings over a 3-week period. This study revealed that of a total of 1455 alarm soundings, only eight signalled potentially life-threatening situations. On the basis of this finding, O’Carroll suggests that a graded system of alarms is required, which is a point that will be discussed in more detail later in this article.

A complementary study\(^8\) again using a retrospective questionnaire method, asked anaesthetists about their alarm use. This study generally showed that alarms were used in a positive and constructive way, but also highlighted a widespread problem; in answer to the question ‘Have you ever deliberately deactivated an audible alarm device at the start of a case?’, the majority replied ‘Yes’. A number of reasons were attributed to this, including the need for peace and quiet, the occurrence of too many false alarms during the task, and the sheer unacceptability of the alarm itself.

These studies highlight a number of interesting psychological points; alarms are confusing and too numerous, they are often non-critical even though they sound as if they are signalling critical events, and they are often switched off because they do not function optimally. Placing the cognitive processes of the user of the system at the centre of the design process leads to a number of alternative design possibilities which should reduce these problems.

**PERSON-CENTRED VERSUS EQUIPMENT-CENTRED ALARMS**

The human operator, in the form of a doctor, nurse, anaesthetist or other specialist, is at the centre of a complex human-machine interface when working within the more technologically advanced medical arenas such as the Intensive Therapy Unit and the Operating Room. It is essential to consider the nature of the tasks placed upon these individuals at a more general level than at the level of a single piece of equipment, especially as there is another human, the patient, at the other end of the system. In the case of alarms and alarm use, there are certain features of the cognitive processing of sound, common to all individuals, which are unlikely to change and which are slowly being incorporated at a design and standards level.

The first of these problems is that there are simply too many alarms, as indicated earlier. How can one get around this problem without creating the potential for missed critical events? Our physiology and anatomy has taken millions of years to evolve, and is not likely to change in the near future; neither is the basic functioning of our cognitive processes, so a useful design procedure might be to focus on these, rather than on technology, which develops and changes dramatically over short periods of time.

Kerr\(^9\) expounds this idea, pointing out that damage to living tissue is caused by relatively few agencies. These categories of risk provide a natural rationale and limit for the number of warnings needed. The categories are hypoxia, ventilation, cardiovascular problems, interruption of artificial perfusion, drug administration, and thermal risk. More specific categories of medical intervention and monitoring can be fitted into these six categories. Thus by this system it is quite possible to monitor all events for which one might wish to be warned, but restricts the number of warnings used to a natural maximum of six or seven sounds.

Although a set of prototype sounds have been designed to support and to demonstrate this principle\(^2\), the sounds proposed are not as important as the concept itself, which is of critical importance because it focuses on the natural capacities and predispositions of the two least flexible features in the system: the limited capacity of human memory and the unchanging physiology (in evolutionary terms) of the patient and the human operator.

A simple analogy between this idea and the designation of football-team colours can be drawn; if one is organizing a tournament with six teams, each of 11 players, then it will be much easier to distinguish between them if each team wears a different colour. It is then immediately clear which teams are currently playing. If we then need more specific information about which player we are looking at, then we can look at the number on the back of his shirt. In the same way, once we have been alerted to the category of alarm, we can go and take a closer look at the patient or the equipment. Just as we cannot distinguish between 66 players that we do not know very well, we cannot distinguish between all of the alarms we are likely to hear. We stand a much better chance of identifying a single alarm if we help ourselves by categorizing the types of risk and cutting down the number of sounds used.

Kerr’s proposal contains one other central theme, and that is that each of the sounds is available in two versions; an emergency and a cautionary version. Medical events and procedures might be ascribed either to the cautionary or to the emergency category. This idea could be expanded in medical monitoring to cover a range of urgency levels.

One of the reasons why false alarms are such a problem is that the normal operating band of alarms is too narrow; one limit is too high, and the other is too low, leading to unnecessary numbers of false alarms. There are other factors such
as the patient moving, which will sometimes activate an alarm, which are unavoidable artefacts. The setting of low alarm points may be done for the purposes of safety, because it is generally thought better to err on the side of caution. Thus the vast majority of alarms signify events that are, at least at that point, non-life-threatening. A graded series of alarms whereby relatively non-urgent situations are signalled by relatively non-urgent alarms would be much more ergonomic, and is achievable. The monitoring of the medical changes themselves, which would ultimately determine the urgency of the situation, is a multi-parameter and complex task which can increasingly be carried out by expert systems, although the critical values themselves may need to be set manually. Advances in monitoring systems, and the assessment of the criticality of medical situations depending upon medical parameter values are, however, not generally supported by advances in the design of the actual alarm sounds which accompany such changes. Nevertheless it is possible, with a graded system of alarms, to increase their normal operating band so as to provide auditory feedback on the current criticality of the situation. Such alarms could easily be less intrusive than they are at present.

EFFECTIVE ALARM SOUNDS

The question as to which sounds are most appropriate for use as alarms and/or monitoring sounds is also of some importance. A set of guidelines for auditory warning design proposes a method of alarm design which makes alarms less aversive and more readily learnt. This method of design also lends itself to the manipulation of urgency, which is important in new standards and in the development of new warnings and monitoring sounds. Figure 1 demonstrates the proposed method of construction. Initially, a short pulse of sound is designed which gives the listener the essential acoustic information required. This should contain several harmonics, which makes the sound easier to localize (many existing alarms consist of only a single tone which makes it impossible to localize, and thus further adds to the identification problem). This pulse of sound can be repeated several times, perhaps at different pitches, and with time intervals between them, in order to construct an alarm which can sound something like an atonal melody with a rhythm. This burst of sound would typically last for from 1 to 2 s. If required, a more lengthy repeated cycle of burst followed by silence would then ensue, until the situation is remedied. Figure 1 represents a prototype and, although variations are possible, especially in the overall cycle of the warning, this principle has now been applied to new and upcoming standards on hospital alarms. Specifically, this principle has been applied to different degrees in an upcoming European standard on electrically-generated alarm signals (prEN 475) and to a worldwide standard on anaesthesia and respiratory care alarm signals (ISO/DIS 9709-2). It is also the basis of a British draft standard currently under development by British Standards Institute Health Centre Committee 16.

Stanford et al. propose a somewhat different, possibly more radical, method of alarm design in which vowel sounds figure centrally. Their studies show that these types of sound are at least more aesthetically acceptable than many currently in use, and so provide an alternative design protocol. If the number of alarms used is dramatically reduced using a logical rationale like the one described above, then the level of confusion between sounds will also dramatically reduce almost automatically. However, even with reduced numbers of alarms there are ways of minimizing warning confusion by designing warnings on a scientific basis. For example, there is a body of research in the cognitive psychology of sound which shows which sorts of aspects of sounds will be most conducive to confusion, and should therefore be avoided (examples include and ). To return to the football analogy, a set of even six or seven colours stands a better chance of being discriminated one from another if the colours are strikingly different; the advantage is somewhat lost if two of the teams are dressed in shades of red and all the others in shades of blue. Unfortunately there seems to be no need to exist between sounds in order to achieve maximal discriminability are not always as obvious as they might seem.

One of the most common types of alarm is that of the continuous tone; that is, a constant pitch which continues to sound until it is turned off. Not only is this bad as an attention-getting device – our perceptual system is geared towards change, rather than constancy – it is bad from a cognitive point of view. It is well known that memory for pitch decays over time and that very few of us are blessed with the ability to recognize absolute pitches. However, many manufacturers still make equipment which produces alarms of this type. A recent study shows that alarms of this type...
sort differing by one and a half octaves (a huge difference in perceptual terms) are often confused with one another, even though they signify quite different medical events.

Temporal patterning is also very important in alarm confusion. Patterson\(^ {11} \) showed that alarms which are spectrally quite different from one another (their sound 'quality' is quite different) were nevertheless confused because they repeated their on/off cycle at the same rate. Meredith and Edworthy\(^ {6} \) demonstrate that warnings which do not share the same repetition rate, but merely share the same proportion of on/off time, are also readily confused.

Aside from the topic of confusion, the concept of urgency is also a central one in the development of new alarm sounds and standards. Although it is not necessary to startle either staff or patients, as many existing alarms often do, it is necessary to attach an appropriate sense of urgency to the alarms which will be used. This can be achieved by a somewhat more musical approach to alarm design. Any graded system of alarms incorporates the tacit assumption that more urgent events will be signalled by more urgent-sounding alarms. The system proposed by Schreiber and Schreiber\(^ {5} \) contains a very simple system of this type, where there are three types of alarm sound, depending upon the urgency of the risk. Here, the differentiation between the urgency levels appear to be based on the temporal features, particularly the length, of the warning signal. For the top priority situations a continuously repeating sound pattern is used, for second priority situations an intermittently repeating sound pattern is used, and for the third priority a single short tone, or no sound at all, is used.

Much greater subtlety in the manipulation of perceived urgency is possible, however, than either the ideas of Schreiber and Schreiber\(^ {5} \) or those of Kerr\(^ {7} \) suggest. With computer technology it is possible to produce a chip very cheaply which will play an acoustically complex sound with many dimensions contributing to its final form. A vast range of urgency can be covered with such a degree of flexibility. Research has been carried out which shows that many dimensions of an alarm, such as its speed and repetition rate, its pitch, its harmonic quality, its melodic pattern and so on affect its perceived urgency\(^ {13} \). More recent research also shows that the effects of the more important parameters such as speed can be quantified and applied to warning design\(^ {14} \). The efficacy of this work has been demonstrated in the design of a prototype set of 13 warnings in which the urgency was predicted prior to testing: experimental results supported the predictions such that the Spearman rank correlation between the predicted urgency order and the experimentally obtained urgency order was highly significant (0.929, \( P < 0.0001 \)). This study used a multiple comparison procedure in which the 13 warnings were systematically compared in groups of four, until all 13 had been compared twice with each other in a block-randomized fashion. Subjects were also very consistent in their rankings, producing a value for Kendall's \( W \), a statistical measure of consistency, of 0.920 which is also highly significant (\( F(11, 154) = 148.6, P < 0.0001 \)). These warnings varied from one another along many dimensions, including their pitch, speed, harmonic content, melodic pattern and so on and were thus more acoustically complex, and therefore distinguishable from one another, than typical warnings. Under these circumstances, many more levels or urgency are possible than most current applications could produce, and the most appropriate application of this work is in the design of sounds to accompany slowly developing medical events in which the level of criticality is changing. Not only would the use of such a graded system help in the auditory monitoring of events, it would also help to overcome the problem of alarms being turned off because they have only one, possibly overly-urgent, form. The relationship between medical parameter values and sound parameter values remains to be elucidated, however, in future research and development work. For example, the potentially developing medical pathology would need to be hierarchically defined in such a way as to allow matching to a systematically graded alarm system.

STANDARDIZATION

At present, the developing European horizontal standard (prEN 475) on electrically-generated alarm signals and the proposed worldwide standard (ISO/DIS 9703-2) on anaesthesia and respiratory care alarm signals are both working towards adopting the idea of having simply a general alarm with two or three graded levels of urgency. This general alarm would exist alongside all the other alarms currently in use. These alarms are specified in the standards in such a way as to make the alarm recognizable from hospital to hospital, although the sound itself can vary to some degree depending on how the manufacturer wishes to construct it. The principles of perceived urgency in sound are applied in these standards, so that the high priority alarm is more urgent than the medium priority (neither are 'alarming' in the traditional way) but there is currently no plan to widen the scope to specific risk categories, as suggested by Kerr\(^ {7} \).

The British standard is, however, proceeding along the lines suggested by Kerr where there are six categories of risk, plus a general alarm, each with two versions of the sound, cautionary and emergency. The cautionary and the emergency forms are clearly distinguishable from one another whilst at the same time being obviously versions of the same sound. The cautionary version is generally slower, quieter, and at a lower pitch than the emergency version and there are other, more subtle, differences between the two.

All three standards represent a step in the right direction in terms of providing a more ergonomic and appropriate user interface; however, there is still scope in both research and development for the match between cognitive processes and alarm use to be improved.
REFERENCES


